

STOCHASTIC ESTIMATION OF REDUCED BASIS ERRORS BASED ON MACHINE LEARNING CONCEPTS

K. Carlberg M. Drohmann

Sandia National Laboratories, Livermore, CA, (mdrohma@sandia.gov)

ModRed CIRM, Marseille, France, June 2013

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration

under contract DE-AC04-94AL85000

ERROR ESTIMATES FOR MODEL ORDER REDUCTION METHODS

Model reduction implies error to high fidelity model: $e := u_h - u_{\text{red}}$.

ERROR BOUNDS

Error bounds η should be

- **rigorous:** $\eta \geq \|e\|$,
- **effective:** $\frac{\eta}{\|e\|} \leq C_{\text{eff}}$ or “not too big”
and
- **efficient**, i.e. quickly computable (as fast as reduced model)

ERROR ESTIMATES FOR MODEL ORDER REDUCTION METHODS

Model reduction implies error to high fidelity model: $e := u_h - u_{\text{red}}$.

ERROR BOUNDS

Error bounds η should be

- **rigorous**: $\eta \geq \|e\|$,
- **effective**: $\frac{\eta}{\|e\|} \leq C_{\text{eff}}$ or “not too big”
and
- **efficient**, i.e. quickly computable (as fast as reduced model)

Often there is a trade-off between **efficiency** and **rigor/effectivity**.

EFFECTIVITY OF ERROR ESTIMATES

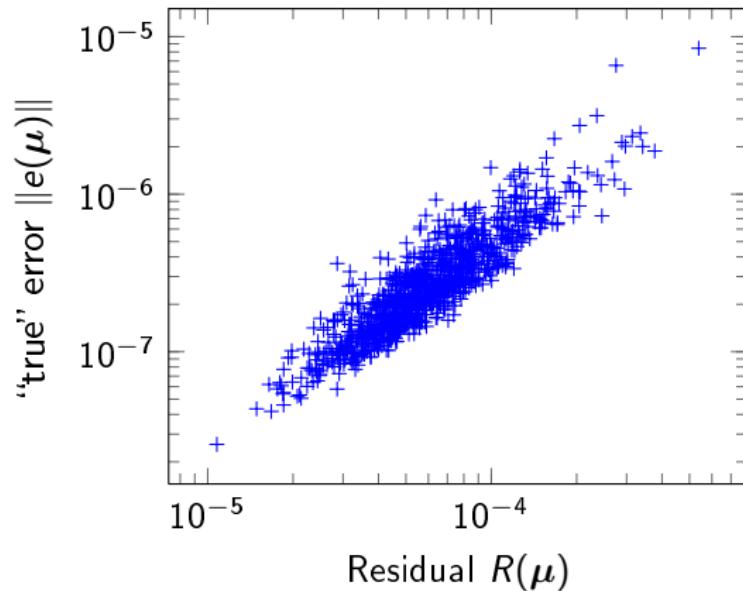
HIGH EFFECTIVITY . . .

- . . . is necessary for the **selection of good** reduced bases.
- . . . ensures **low** reduced basis size for (certified) reduced solutions.

BUT . . .

Effective error estimates can be very **expensive**.

OBSERVATION: STRUCTURED CORRELATION BETWEEN RESIDUALS AND ERRORS



- **Goal:** Use structure for an **efficient** and **effective** stochastic error estimate.
- In UQ context **rigor** is often not important.

OUTLINE

① A POSTERIORI ERROR ESTIMATORS

- Reduced basis scheme
- Ingredients
- Effectivity of error estimators

② STOCHASTIC ERROR ESTIMATE

③ NUMERICAL EXPERIMENTS

- Thermalblock-Problem

④ SUMMARY AND FUTURE WORK

OUTLINE

① A POSTERIORI ERROR ESTIMATORS

- Reduced basis scheme
- Ingredients
- Effectivity of error estimators

② STOCHASTIC ERROR ESTIMATE

③ NUMERICAL EXPERIMENTS

- Thermalblock-Problem

④ SUMMARY AND FUTURE WORK

HIGH-DIMENSIONAL MODEL

$$(\partial_t u_h) + H \left\{ \begin{array}{c} \mathcal{L}_h \\ \text{---} \\ \text{---} \end{array} \right\} = u_h = b_h$$



OFFLINE PHASE

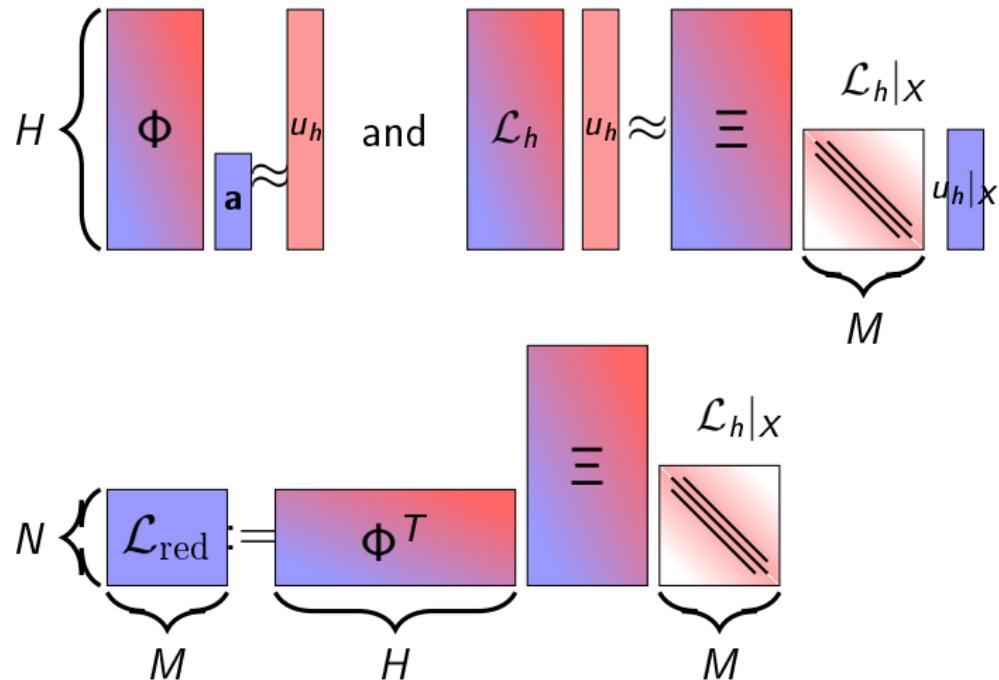
- Find reduced basis functions $\Phi := [\varphi_1 \dots \varphi_N]$,
- such that

$$H \left\{ \begin{array}{c} \Phi \\ \text{---} \\ a \\ \text{---} \\ u_h \end{array} \right. \quad \text{and}$$

$$N \left\{ \begin{array}{c} \mathcal{L}_{\text{red}} \\ \text{---} \\ N \end{array} \right. := \underbrace{\Phi^T}_{H} \underbrace{\mathcal{L}_h}_{H} \underbrace{\Phi}_{N}$$

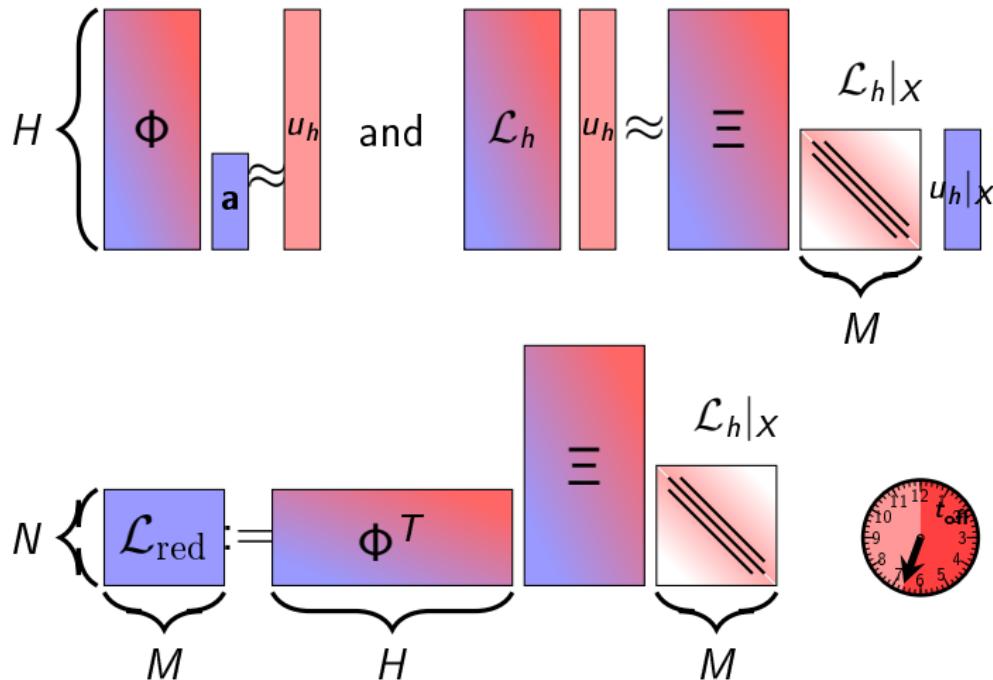
OFFLINE PHASE

- Find reduced basis functions $\Phi := [\varphi_1 \dots \varphi_N]$,
- interpolation points $X := \{x_1, \dots, x_M\}$ and functions $\Upsilon := [\xi_1 \dots \xi_M]$,
- such that



OFFLINE PHASE

- Find reduced basis functions $\Phi := [\varphi_1 \dots \varphi_N]$,
- interpolation points $X := \{x_1, \dots, x_M\}$ and functions $\Upsilon := [\xi_1 \dots \xi_M]$,
- such that



REDUCED BASIS MODEL

$$(\partial_t u_{\text{red}} +) \quad N \quad \left\{ \begin{array}{c} \mathcal{L}_{\text{red}} \\ \Phi|x \end{array} \right. \quad \left. \begin{array}{c} \mathbf{a} \\ = b_{\text{red}} \end{array} \right.$$

$M \qquad N$



Error: $e := u_h - u_{\text{red}}$ with $u_{\text{red}} := \Phi \mathbf{a}$

A POSTERIORI ERROR ESTIMATOR (DHO, 2012)

ESTIMATOR

$$\|u_h^k(\boldsymbol{\mu}) - u_{\text{red}}^k(\boldsymbol{\mu})\| \leq \eta_{N,M,M'}^k(\boldsymbol{\mu})$$

A POSTERIORI ERROR ESTIMATOR

THEOREM (A POSTERIORI ERROR ESTIMATOR)

Assumptions:

- Operator(s) fulfill "Lipschitz" properties:
 - ▶ $\|u - v + \Delta t \mathcal{L}_I[u] - \Delta t \mathcal{L}_I[v]\| \geq \frac{1}{c_{I,\Delta t}} \|u - v\|_{W_h}$
 - ▶ $\|u - v - \Delta t \mathcal{L}_E[u] + \Delta t \mathcal{L}_E[v]\| \leq C_{E,\Delta t} \|u - v\|_{W_h}$
- M' -trick: Empirical interpolations exact for larger number of interpolation points $M + M'$ and the initial data is $u_0(\mu) \in \text{span}\Phi$

A POSTERIORI ERROR ESTIMATOR

THEOREM (A POSTERIORI ERROR ESTIMATOR CONT.)

Assumptions:

- Operator(s) fulfill “Lipschitz” properties:
 - ▶ $\|u - v + \Delta t \mathcal{L}_I[u] - \Delta t \mathcal{L}_I[v]\| \geq \frac{1}{C_{I,\Delta t}} \|u - v\|_{W_h}$
 - ▶ $\|u - v - \Delta t \mathcal{L}_E[u] + \Delta t \mathcal{L}_E[v]\| \leq C_{E,\Delta t} \|u - v\|_{W_h}$
- M' -trick: Empirical interpolations exact for larger number of interpolation points $M + M'$ and the initial data is $u_0(\mu) \in \text{span}\Phi$

Then:

$$\|u_{\text{red}}^k(\mu) - u_h^k(\mu)\| \leq \eta_{N,M,M'}^k(\mu)$$

with

$$\eta_{N,M,M'}^k(\mu) := \sum_{i=0}^{k-1} C_{I,\Delta t}^{k-i+1} C_{E,\Delta t}^{k-i} \left(\|R_{I+E,M,M'}^{k+1}(\mu)\| + \|\Delta t R^{k+1}(\mu)\| + \varepsilon^{New} \right)$$

A POSTERIORI ERROR ESTIMATOR

THEOREM (A POSTERIORI ERROR ESTIMATOR CONT.)

Then:

$$\|u_{\text{red}}^k(\mu) - u_h^k(\mu)\| \leq \eta_{N,M,M'}^k(\mu)$$

with

$$\eta_{N,M,M'}^k(\mu) := \sum_{i=0}^{k-1} C_{I,\Delta t}^{k-i+1} C_{E,\Delta t}^{k-i} \left(\|R_{I+E,M,M'}^{k+1}(\mu)\| + \|\Delta t R^{k+1}(\mu)\| + \varepsilon^{New} \right)$$

The residuals $R_{*,M}$ measure the empirical interpolation error, e.g.

$$R_{*,M,M'}^{k+1,\nu} := \sum_{m=M}^{M+M'} I_m^* \left[u_{\text{red}}^{k+1,\nu} \right] \xi_m = \mathcal{I}_{M+M'}[\mathcal{L}] - \mathcal{I}_M[\mathcal{L}]$$

A POSTERIORI ERROR ESTIMATOR

THEOREM (A POSTERIORI ERROR ESTIMATOR CONT.)

Then:

$$\|u_{\text{red}}^k(\mu) - u_h^k(\mu)\| \leq \eta_{N,M,M'}^k(\mu)$$

with

$$\eta_{N,M,M'}^k(\mu) := \sum_{i=0}^{k-1} C_{I,\Delta t}^{k-i+1} C_{E,\Delta t}^{k-i} \left(\|R_{I+E,M,M'}^{k+1}(\mu)\| + \|\Delta t R^{k+1}(\mu)\| + \varepsilon^{\text{New}} \right)$$

empirical interpolation error

Galerkin projection error

time evolution

Newton step error

A POSTERIORI ERROR ESTIMATOR

THEOREM (A POSTERIORI ERROR ESTIMATOR CONT.)

Then:

$$\|u_{\text{red}}^k(\mu) - u_h^k(\mu)\| \leq \eta_{N,M,M'}^k(\mu)$$

with

$$\eta_{N,M,M'}^k(\mu) := \sum_{i=0}^{k-1} C_{I,\Delta t}^{k-i+1} C_{E,\Delta t}^{k-i} \left(\|R_{I+E,M,M'}^{k+1}(\mu)\| + \|\Delta t R^{k+1}(\mu)\| + \varepsilon^{\text{New}} \right)$$

empirical interpolation error	(quasi-) rigorous, M' affects efficiency and effectivity
Galerkin projection error	rigorous and efficient
time evolution	selection affects effectivity
Newton step error	neglectable

EFFECTIVE ERROR BOUND

- (Compute many interpolation points M' .)
- Define Lipschitz-constants as tight as possible by approximation of $C_{I,\Delta t} = \sup_{k=0,\dots,K} \left\| \left(D(\text{Id} + \mathcal{L}_{I,\Delta t})|_{u_h^k} \right)^{-1} \right\|$, e.g. as

COMPUTATION OF LIPSCHITZ CONSTANTS

- Operator-norm approximation:

$$C_{I,\Delta t} \approx \inf_{v \in X_{\text{test}}} \frac{\left\| D(\text{Id} + \mathcal{L}_{I,\Delta t})|_{u_h^k}[v] \right\|}{\|v\|}$$
$$C_{E,\Delta t} \approx \sup_{v \in X_{\text{test}}} \frac{\left\| D(\text{Id} + \mathcal{L}_{E,\Delta t})|_{u_h^k}[v] \right\|}{\|v\|},$$

where X_{test} is a set of intermediate Newton step solutions.

- $C_{I,\text{low}} \approx C_I(\mu)$, $C_{E,\text{upper}} \approx C_E(\mu)$ for all $\mu \in \mathcal{P}$

REMARK: OUTPUT OF INTERESTS

- If application has output of interest: $s(u_h)$ (s linear functional),
- the error bound $\eta^s \geq |s(u_h) - s(u_{\text{red}})|$
- can be improved by solving bounds for adjoint problems.

ELLIPTIC PROBLEMS

$$\begin{aligned} A(\mu) [u_h(\mu), \phi] &= b(\phi) && \text{for all } \phi \in \mathcal{W}_h \\ A(\mu) [u_{\text{red}}(\mu), \phi] &= b(\phi) && \text{for all } \phi \in \mathcal{W}_{\text{red}} \end{aligned}$$

ELLIPTIC PROBLEMS

$$\begin{aligned} A(\mu) [u_h(\mu), \phi] &= b(\phi) && \text{for all } \phi \in \mathcal{W}_h \\ A(\mu) [u_{\text{red}}(\mu), \phi] &= b(\phi) && \text{for all } \phi \in \mathcal{W}_{\text{red}} \end{aligned}$$

ERROR ESTIMATOR

$$\eta(\mu) := \frac{R(\mu)}{\beta(\mu)}$$

with

- residual $R(\mu) := \sup_{\phi \in \mathcal{W}_h} A(\mu) [u_{\text{red}}(\mu)] - b(\phi)$ and
- coercivity constant $\beta(\mu)$.

EFFECTIVITY OF ERROR ESTIMATORS

FACTORS

- Coercivity / Lipschitz constants
- Propagation over time
- Output functional

SOLUTIONS FROM LITERATURE:

<i>Method</i>	<i>strategy</i>	<i>cost</i>
SCM	finds tight coercivity constant	increased basis size / offline time
EV approximation	compute approximation of eigenvalues from reduced matrices	increased basis size / offline time
Space-time grid	tighter control of error propagation over time	rewrite of the numerical scheme

OUTLINE

① A POSTERIORI ERROR ESTIMATORS

- Reduced basis scheme
- Ingredients
- Effectivity of error estimators

② STOCHASTIC ERROR ESTIMATE

③ NUMERICAL EXPERIMENTS

- Thermalblock-Problem

④ SUMMARY AND FUTURE WORK

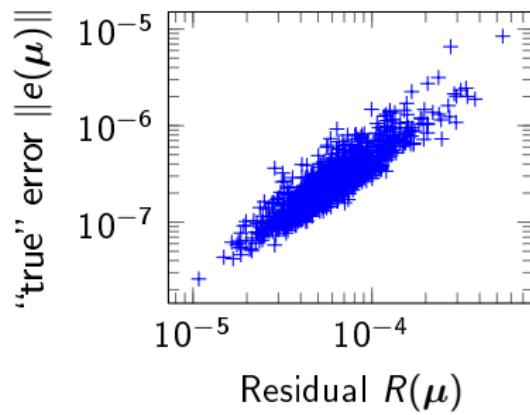
MACHINE LEARNING SOLUTION

OBSERVATION

Tuple set of residuals and errors

$$F : \mu \mapsto [R(\mu), e(\mu)]$$

usually behaves very smoothly.



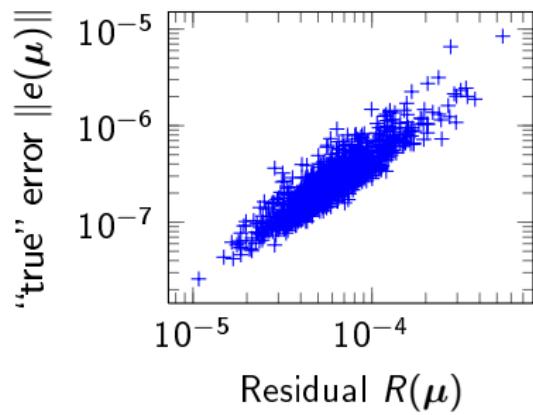
MACHINE LEARNING SOLUTION

OBSERVATION

Tuple set of residuals and errors

$$F : \mu \mapsto [(X(\mu) := (R(\mu), J(\mu))) , e(\mu)]$$

usually behaves very smoothly, for indicators $J(\mu)$.



MACHINE LEARNING SOLUTION (CONT.'D)

- *Assumption:* Tuples $(X(\mu), e(\mu))$ have an underlying “noisy” structure.
- Find stochastic process \bar{f} such that

$$\bar{f}(x) \sim \{e \mid \exists \mu \in \mathcal{P} \text{ with } f(\mu) = (x, e)\}.$$

MACHINE LEARNING SOLUTION (CONT.'D)

- *Assumption:* Tuples $(X(\mu), e(\mu))$ have an underlying “noisy” structure.
- Find stochastic process \bar{f} such that

$$\bar{f}(x) \sim \{e \mid \exists \mu \in \mathcal{P} \text{ with } f(\mu) = (x, e)\}.$$

- First try: Assume **Gaussian process**

$$\bar{f}(x) \sim \mathcal{N}(m(x), \Sigma).$$

CONSTRUCTION OF GAUSSIAN PROCESS

1. Compute training samples $(X(\mu), e(\mu))$ for $\mu \in P_{\text{train}} \subset \mathcal{P}$
2. Infer Gaussian process predictions $\bar{f}(x^*)$ at new points x^* with
 - A. GP Kernel methods (Rasmussen&Williams, 2005)
 - B. Relevance vector machine (Tipping, 2001)

GP KERNEL METHOD

ASSUMPTION

$$\mathbf{x} \in \mathbb{R}^N, \quad \text{cov}(\bar{f}(\mathbf{x})) = K(\mathbf{x}) + \sigma^2 I$$

with **kernel**

$$(K(\mathbf{x}))_{1 \leq n, m \leq N} = \exp\left(\frac{(\mathbf{x}_n - \mathbf{x}_m)^2}{r^2}\right).$$

GP KERNEL METHOD

ASSUMPTION

$$\mathbf{x} \in \mathbb{R}^N, \quad \text{cov}(\bar{f}(\mathbf{x})) = K(\mathbf{x}) + \sigma^2 I$$

with **kernel**

$$(K(\mathbf{x}))_{1 \leq n, m \leq N} = \exp\left(\frac{(\mathbf{x}_n - \mathbf{x}_m)^2}{r^2}\right).$$

1. Infer **hyper-parameters** σ^2 and r .
2. **Predict** $\bar{f}(x^*) = \mathcal{N}(m(x^*), \Sigma)$ for test/validation values x^* .

RELEVANCE VECTOR MACHINE

ASSUMPTION

$$\bar{f}(x) = \sum_{m=1}^M w_m \phi_m(x) + \epsilon$$

with fixed basis functions ϕ_m (e.g. polynomials), priors $w \sim \mathcal{N}(0, A)$ and noise $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$.

RELEVANCE VECTOR MACHINE

ASSUMPTION

$$\bar{f}(x) = \sum_{m=1}^M w_m \phi_m(x) + \epsilon$$

with fixed basis functions ϕ_m (e.g. polynomials), priors $w \sim \mathcal{N}(0, A)$ and noise $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$.

1. Infer **hyper-parameters** σ^2 and A .
2. **Predict** $\bar{f}(x^*) = \mathcal{N}(m(x^*), \Sigma)$ for test/validation values x^* .

OUTLINE

① A POSTERIORI ERROR ESTIMATORS

- Reduced basis scheme
- Ingredients
- Effectivity of error estimators

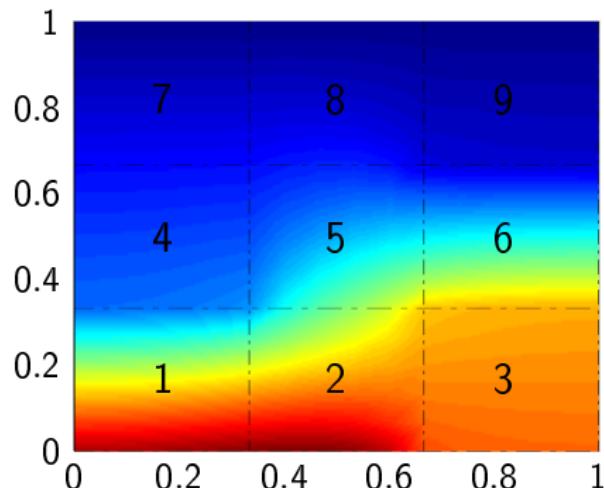
② STOCHASTIC ERROR ESTIMATE

③ NUMERICAL EXPERIMENTS

- Thermalblock-Problem

④ SUMMARY AND FUTURE WORK

EXAMPLE: A THERMAL BLOCK

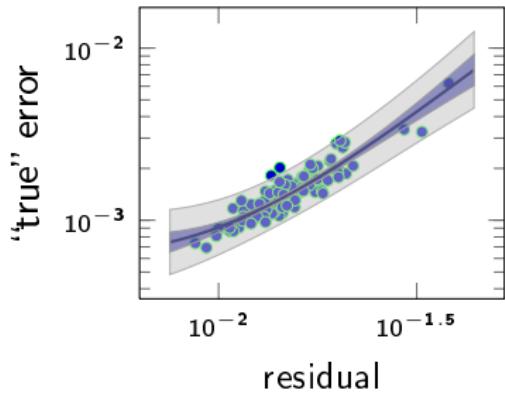


THERMAL BLOCK

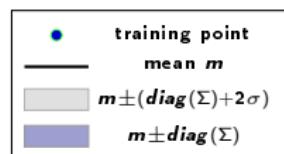
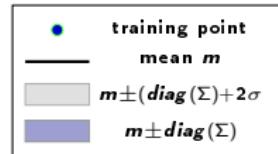
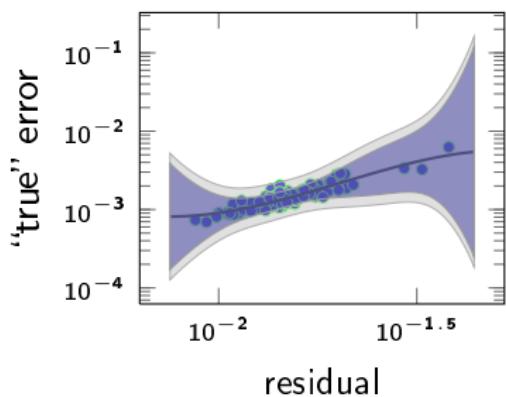
- Heat equation $-\mathbf{a}(\mathbf{x}; \mu) \Delta \mathbf{u} = \mathbf{f}$
- with $\mathbf{a}(\mathbf{x}; \mu) = \sum_{i=1}^9 \mu_i \chi_{\Omega_i}(\mathbf{x})$
- 9 parameters

VISUALIZATION OF GAUSSIAN PROCESSES

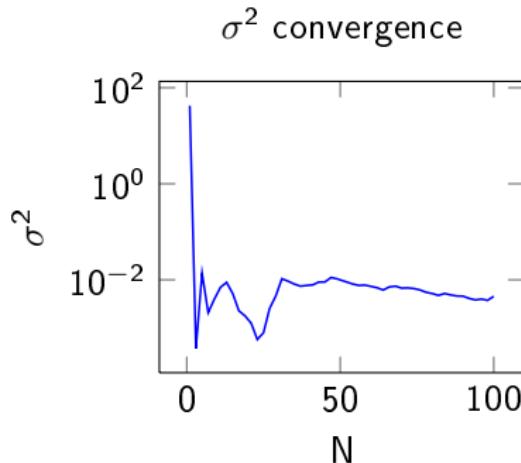
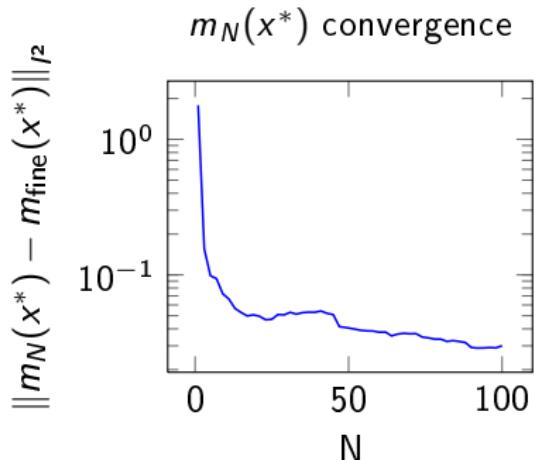
GP-Kernel method



RVM method (polynomial basis)

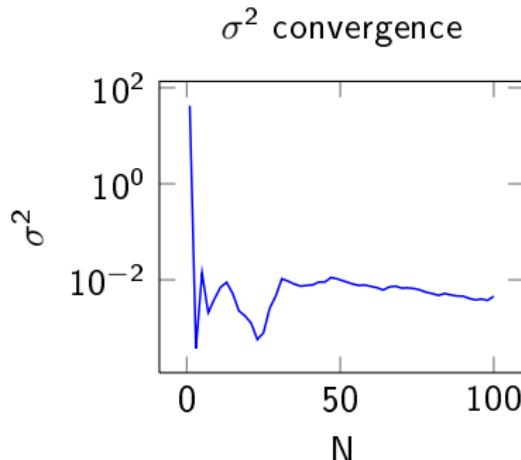
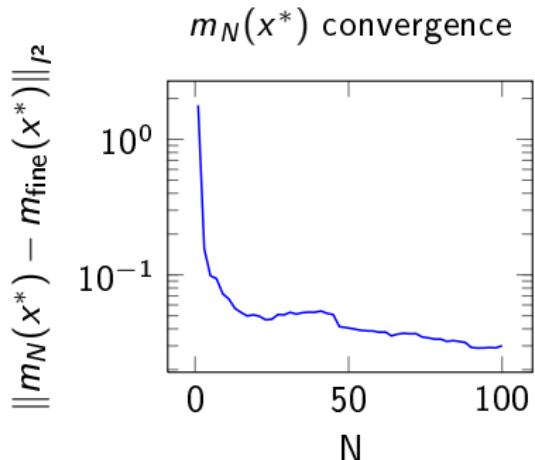


CONVERGENCE OF $m_N(x)$ AND σ_N^2



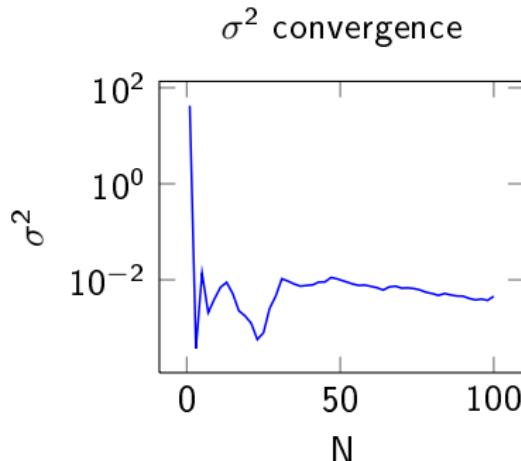
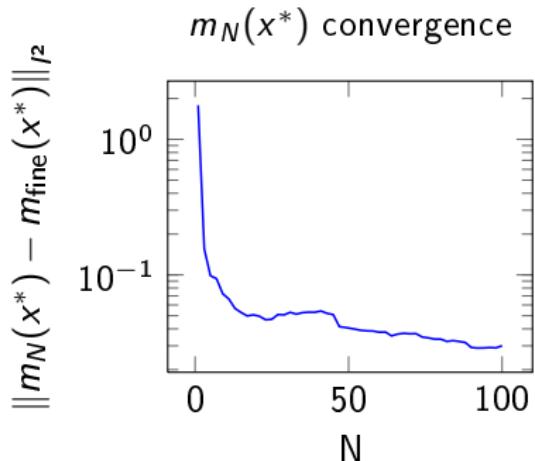
- validation sample size = 200
- $N_{\text{fine}} = 1000$

CONVERGENCE OF $m_N(x)$ AND σ_N^2



- validation sample size = 200
- $N_{\text{fine}} = 1000$
- Stabilizes after 30–40 training samples.
- The estimator seems to converge.

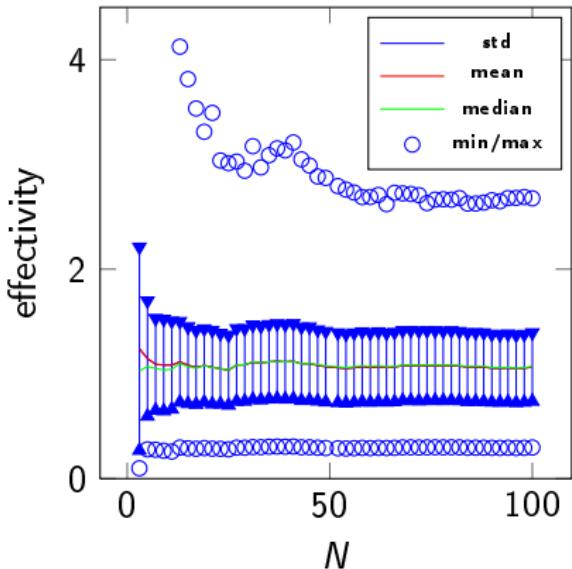
CONVERGENCE OF $m_N(x)$ AND σ_N^2



- validation sample size = 200
- $N_{\text{fine}} = 1000$
- Stabilizes after 30–40 training samples.
- The estimator seems to converge.
- Structural assumption is **validated**.

EFFECTIVITY

Effectivity analysis over 200 validations



-	m_{47}	m_{100}	η
mean	1.08	1.07	424.41
median	1.09	1.07	256.41
std	0.32	0.31	433.72
min	0.3	0.3	68.25
max	2.89	2.68	3,550.8

Stochastic error estimates

- are more effective (400x), but
- underestimate the error sometimes.

OUTLINE

① A POSTERIORI ERROR ESTIMATORS

- Reduced basis scheme
- Ingredients
- Effectivity of error estimators

② STOCHASTIC ERROR ESTIMATE

③ NUMERICAL EXPERIMENTS

- Thermalblock-Problem

④ SUMMARY AND FUTURE WORK

SUMMARY AND FUTURE WORK

SUMMARY

- Based on **Bayesian inference**.
- **Effective** a posteriori error bounds are **costly**.
- Given a sufficient number of high dimensional computations, **cheap** and **effective** error bounds can be constructed.

SUMMARY AND FUTURE WORK

SUMMARY

- Based on **Bayesian inference**.
- **Effective** a posteriori error bounds are **costly**.
- Given a sufficient number of high dimensional computations, **cheap** and **effective** error bounds can be constructed.

FUTURE WORK

- Use (non-Gaussian) processes in order to reduce likelihood for underestimating.
- Non-linear and time dependent examples with more complex indicators.
- Try to improve effectivity with further indicators (e.g. eigenvalues of reduced matrices).